**Rhonald Reside  
DS-670  
Assignment 10**

**Draft Manuscript**

**Results**

The data files that the weather data set came in was JSON file format. Weather data set came with seven different types of files (i.e., variable). It came in with Dew Point, which was in degrees Celsius. Humidity was the next variable which came in percentage. Pressure was the next variable which came in the measurement of mBar. Temperature was the next variable which was measured in degrees Celsius. Wind direction was the next set of variables which was came in the measurement of degrees. The next variable is wind speed, which was measured in kilometers per hour. The final variable was visibility.

I discovered that there was a data gap somewhere in the month of June. I found that there was a data gap in from June 8, 2014 to August 1, 2014. I could see this once I created a line graph with temperature, dew point and humidity. It was a very interesting visual find. I suspect that since this this data gap occurred throughout all the variable, there was most likely the possibility of data storage issue or they all ran out of battery at the same time.

Used Tableau to visualize data set. I find it easiest to visualize my data after it was loaded into Tableau. I created line graphs, heat maps and averages to help me build a story from the data. I discovered a few things after this step using Tableau. I found that there was a data gap in from June 8, 2014 to August 1, 2014. I could see this once I created a line graph with temperature, dew point and humidity. It was a very interesting visual find. I suspect that since this this data gap occurred throughout all the variable, there was most likely the possibility of data storage issue or they all ran out of battery at the same time.

We can visually see that there was a relationship between Temperature, Dew point and Humidity. From the graph, we discover a few other things. The next discovery I found was the warmest day and the coldest day in the data set. I found that August 2, 2014 was the warmest day measured at 80.60° F with a humidity of 37 and a dew point of 11. The coldest day occurred March 11, 2014 with a humidity of 84 and a dew point of -4. I decided to create a heat map that represented the high temperature all the days. I also did the same for the cold and created heat map of the coldest temperature day.

I split my data into two data sets. I created a training data set and a test data set. This is an important step because you can run your statistics on your training data set. I used I plan to use temperature as my Y-Variable, also known as my dependent variable. I label my variables as dew point and humidity; we would also know this as independent variables. The glm() function will create a confusion matrix. A confusion matrix is a table used to describe the classification performance of my model of my training data set. In this case, it shows us the accuracy of our classifier of 65°.

In Tableau, I created a heat map for hot and cold days. I found was the warmest day and the coldest day in the data set. I found that August 2, 2014 was the warmest day measured at 80.60° F with a humidity of 37 and a dew point of 11. The coldest day occurred March 11, 2014 with a humidity of 84 and a dew point of -4. I decided to create a heat map that represented the high temperature all the days. I also did the same for the cold and created heat map of the coldest temperature day.

I used Zeppelin to further analyze my weather data. I aggregated the data using Zeppelin to create a subset of warmest and coldest day. One neat feature I was able to create was this drop box. It gave the ability to extract the temperature, dew point (average), and humidity (average). It confirmed the accuracy of Tableau and R programming.

My competitors work produced an error rate of less than 5 percent. He explained and confirmed the Direct relationship between temperature, humidity and dew point. We see my error rate is .09 percent with my classification rate of 90.34 percent. The classification decision was based on a binary response of 65˚. This was my threshold. My logistic regression was able to predict a 90.3% accuracy rate on the test data.

My work did outperform my competitors work because first on speed, he wanted to show that you can calculate without using a computer. My work is from computers and program. There is no way hand calculation can beat a computer and produce my high accuracy rate.

I agreed with my competitor that there was a direct relationship with temperature, humidity and dew point. We saw that in the very beginning with the use of Tableau. We confirmed using the glm() function in R. Our statistics showed the relationship.

We also saw that from the dataset, the warmest day was on August 2, 2014 at 80.6⁰ F, with average humidity of 37 and dew point of 11. We also found Coldest day was on March 11 at 26.6⁰ F, with average humidity of 84 and dew point of -4.